Hot-Carrier Generation in Submicrometer VLSI Environment

TAKAYASU SAKURAI, MEMBER, IEEE, KAZUTAKA NOGAMI, MASAKAZU KAKUMU, AND TETSUYA IIZUKA, MEMBER, IEEE

Abstract—**Submicrometer MOSFET's may suffer from reliability degradation, which has a strong correlation with substrate current. In order to know what is happening to substrate current in a VLSI environment, a substrate-current circuit simulator is developed. The simulator is applied to MOS unit circuit blocks, VLSI static memories, and dynamic memories and their hot-carrier duty ratios are calculated. A new circuit technology called Normally-On Enhancement MOSFET Insertion (NOEMI) is proposed which can suppress hot-carrier generation. Several design implications for submicrometer VLSI's are obtained through the analysis.**

I. Introduction

Reliability degradation of small geometry MOSFET's due to hot carriers is one of the most serious problems that prevail for high-speed and high-density VLSI's. It is known [1] that the threshold voltage shift and conductance degradation have close relation with substrate current which is a good monitor of the amount of hot carriers generated. Much effort has been devoted to characterizing the degradation by means of dc measurements and there has been some work on ac measurement of n-channel inverters [2]. However, MOSFET's in a complex circuit see unique bias conditions, because the drain and gate voltage change simultaneously in time. In order to understand the hot-carrier generation in VLSI environment, a substrate-current circuit simulator is developed and applied to MOS complex circuit blocks and VLSI's [3].

It is found that a serial connection of n-channel transistors is effective in reducing hot-carrier generation. Using this principle, new circuits to suppress the hot-carrier generation are proposed. There are other VLSI design implications obtained by the analysis for static and dynamic memories. Especially in dynamic RAM's, the MOSFET's which discharge bootstrapped nodes generate much more hot carriers than other transistors operated under normal voltage of V_{cc}. A remedy for this situation is Normally-On Enhancement MOSFET Insertion (NOEMI) circuit technology. For dynamic RAM's (DRAM), a half V_{cc} precharge scheme is shown to improve hot-carrier generation compared with a V_{cc} precharge of bit-lines.

Section II of this paper describes a basic numerical model of substrate current and its simple applications. Section III is for new circuit technology which suppress hot-carrier generation. Sections IV and V deal with substrate-current analysis of static and dynamic RAM cases, respectively. Section VI is dedicated to conclusions.

Manuscript received May 1, 1985; revised August 26, 1985.
The authors are with the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, 210 Japan.
IEEE Log Number 8406279.

II. Basic Model and Simple Applications

The substrate-current model is basically a combination of Sing *et al.*'s [4] and Hu's [5], modified to include substrate bias effects in the form of Taylor's expansion. Previous models fail to reproduce back-gate bias effects. The concrete expressions for 0.8- μm poly-gate-length conventional devices with oxide thickness of 16 nm are shown below:

$$I_{sub} = I_{ds} * a(V_{ds} - V_{dsat})^b$$
$$a = 2.24*10^{-5} - 0.10*10^{-5}V_{ds}$$
$$b = 6.4$$
$$V_{dsat} = \frac{V_{gsth} L_{eff} E_{sat}}{V_{gsth} + L_{eff} E_{sat}}$$
$$V_{gsth} = V_{gs} - V_{th} - 0.13V_{bs} - 0.25V_{gs}$$
$$E_{sat} = 1.10*10^7 + 0.25*10^7 V_{gs}$$

where I_{sub} is for substrate current, I_{ds} for drain–source current, V_{ds} for drain–source voltage, V_{gs} for gate–source voltage, V_{bs} for bulk–source voltage, V_{dsat} for drain saturation voltage, L_{eff} for effective channel length in meters, and E_{sat} for velocity saturation field in volts per meter.

A substrate current is basically expressed as a drain–source current multiplied by a factor, the form of which has been introduced by Sing *et al.* somewhat empirically. They concluded that the best value for the parameter b is 7.0. The expression for the drain saturation voltage has been reported by Hu and can be considered as a harmonic average of a conventional $(V_{gs} - V_{th})$ term and a velocity saturation term. Linear modification terms are added empirically to best reproduce the measured results.

The resulting model includes seven empirical parameters, which are given in concrete numbers in the above expressions. I_{ds} is modeled by level 3 model of SPICE [6]. The matching between measured results and the model calculation is good as shown in Fig. 1. Fitting was also successful

0018-9200/86/0200-0187$01.00 ©1986 IEEE

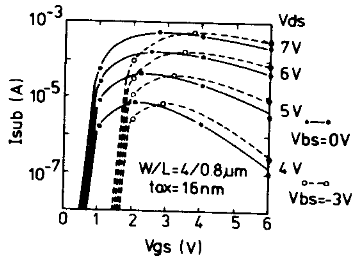


Fig. 1. Numerical model of substrate current. Circles are measured points. This device is the conventional n-channel MOSFET. The fitting was also successful for a 1.0-μm LDD device [3].

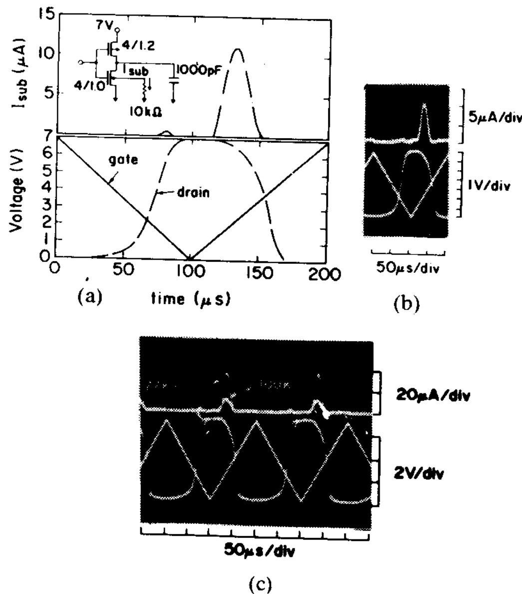


Fig. 2. Substrate current waveforms of a submicrometer inverter. (a) Simulated waveform, (b) measured waveform including charging–discharging current of capacitances which is not hot, and (c) substrate current of CMOS inverter at 77 K and room temperature. The circuit exhibits three times larger substrate current at 77 K, but the circuit behavior is essentially unchanged.

for 1.0-μm poly-gate-length LDD MOSFET if the set of seven parameters is chosen appropriately [3].

The model is implemented in a program written in "awk" of UNIX which serves as a post processor of SPICE. First, SPICE simulates voltage waveforms of all nodes in a circuit. Then, the post processor calculates the substrate-current waveforms using the SPICE output as input.

Fig. 2(a) and (b) is a comparison between simulated and measured substrate-current waveforms for a CMOS inverter consisting of 1-μm MOSFET. The agreement is satisfactory. It should be noted that substrate-current of typically loaded circuit generated at turning-off of the n-channel MOSFET is negligibly small (by over two orders of magnitude) compared with a turning-on case. In Fig. 2(c) substrate current of CMOS inverter at 77 K and room temperature is shown. The circuit exhibits three times larger substrate current at 77 K, but the circuit behavior is essentially unchanged.

Fig. 3 shows a dependence of substrate current of an inverter on load capacitance. Positive dependence of the hot-carrier generation on the load capacitance is observed.

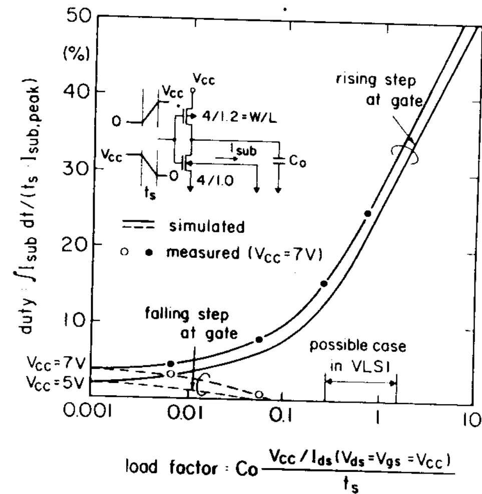


Fig. 3. Substrate current for various load conditions. The heavily loaded inverters, such as bus drivers, are in serious condition in terms of hot-carrier generation. In the measurements, t_s was set to 100 μs and I_{dso} was 1.24 mA. Load capacitances for four measured black circles are 100, 1000, 5000, and 11 000 pF from the left. In the simulation, two values are used for t_s, namely, 100 μs and 1 ns. For t_s of 1 ns, eight orders of magnitude smaller capacitances are used compared with the 100-μs case. The duty was essentially the same between these two cases if the load factors were equal.

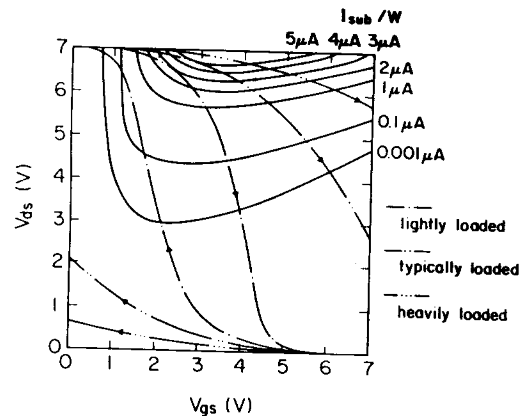


Fig. 4. $V_{gs} - V_{ds}$ trajectories of CMOS inverters and equi-I_{sub} plot. Load capacitance is assumed to be 10, 100, and 1000 fF for lightly, typically, and heavily loaded conditions, respectively. The inverter configuration is the same as that in Fig. 3. t_s is set to 1 ns in the calculation.

It is to be mentioned that lower load capacitance is preferable.

The vertical axis of Fig. 3 is a normalized hot-carrier charge generated in one transient. $I_{sub,peak}$ in the normalizing factor is the maximum dc substrate current of the n-channel MOSFET generated when the drain voltage is set to V_{cc} and the gate voltage is varied. This value of $I_{sub,peak}$ is often employed to characterize the dc hot-carrier degradation of MOSFET's, so that the normalizing factor ($t_s * I_{sub,peak}$) is the hot-carrier charge generated under the severest dc stress. It should be noted that $I_{sub,peak}$ is not a peak value of a substrate-current waveform observed in the time-domain measurements. That is, $I_{sub,peak}$ remains constant if V_{cc} and size and structure of the n-channel transistor are kept unchanged, even if the load capacitance and/or the circuit configuration is changed. The above definition of the $I_{sub,peak}$ is used throughout this paper.

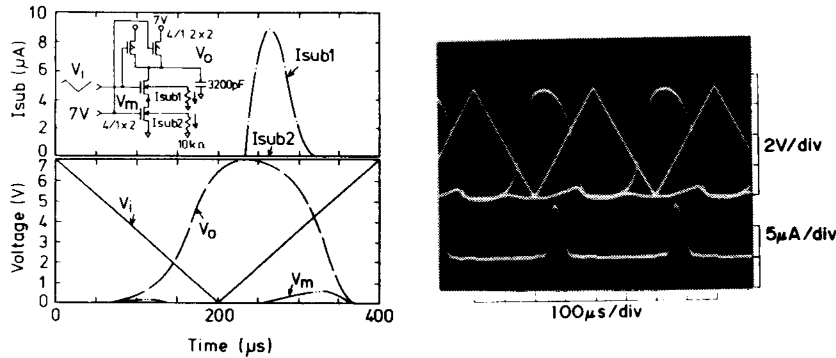


Fig. 5. Comparison of simulated and measured waveforms for NAND gate, when the lower transistor turns on first followed by the upper transistor.

The results of Figs. 2 and 3 can be explained by using $V_{gs}-V_{ds}$ trajectories together with constant substrate-current plot as shown in Fig. 4. More concisely, Fig. 4 is a $V_{gs}-V_{ds}$ plane where constant I_{sub} contours are shown. $V_{gs}-V_{ds}$ bias trajectories felt by the n-channel MOSFET in the inverter are superimposed on this plane. It is seen that the trajectory for heavily loaded inverter passes through the higher I_{sub} region and that the trajectory for discharging case goes through much higher I_{sub} region than charging case.

Another example of the simulator is for NAND circuit when the lower input rises after the upper input rises, where the lower input means the input near to V_{ss} and the upper input signifies the input near to the output node. The comparison of the simulated and measured results is in Fig. 5, which shows good agreement.

III. Hot-Carrier Suppressing Circuit Technology

When the lower input (input *I*2) rises after the upper input (input *I*1), NAND circuits in Fig. 6(a) have better hot-carrier characteristics than inverters and OR gates. Serial connection of nMOSFET's relaxes drain voltage of each MOSFET and hence hot-carrier generation. Fig. 7(a) shows $V_{gs}-V_{ds}$ trajectories in this circumstance. The trajectories do not go through high V_{ds} region. Because of the exponential dependence of substrate current on V_{ds}, the circuit generates more than two orders of magnitude smaller hot carriers compared with an inverter as shown in Fig. 7(b). In this sense, a NAND-rich scheme is recommended for logic VLSI's if careful design considerations can be made so that the upper n-channel transistor turns on first followed by the lower transistor. In the same sense, a Clocked CMOS [7] (C^2MOS) in Fig. 6(b), where a clock comes after input signals settle, generates less hot carriers than a transmission gate in Fig. 6(c) by a factor of a hundred.

Using the same principle, HOt-carrier REsistant Logic (HOREL) is proposed, which can reduce the generated hot carriers by more than two orders of magnitude. The salient feature of the HOREL is NOEMI at the top of the n-channel logic like in Fig. 8. The HOREL at 5 V operates faster than that of normal inverters at 3 V by a factor of 10–40 percent, depending on the logic configuration, while maintaining the hot-carrier generation at the same order. This speed advantage comes from the following reasons. In the HOREL, p-channel logic part is operated under a 5-V

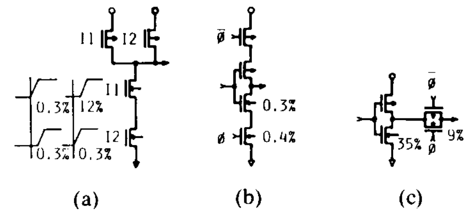


Fig. 6. Calculated duty ratios in typical (a) NAND gate, (b) C^2MOS gate, and (c) transmission gate. The definition of the duty ratio for this figure is the same as that in Fig. 3. The dimension of p-channel and n-channel transistors are as follows: $Wp/Lp = 4\ \mu m/1.2\ \mu m$, $Wn/Ln = 4\ \mu m/1.0\ \mu m$. t_s is 0.3 ns and load capacitance is 40 fF for all cases.

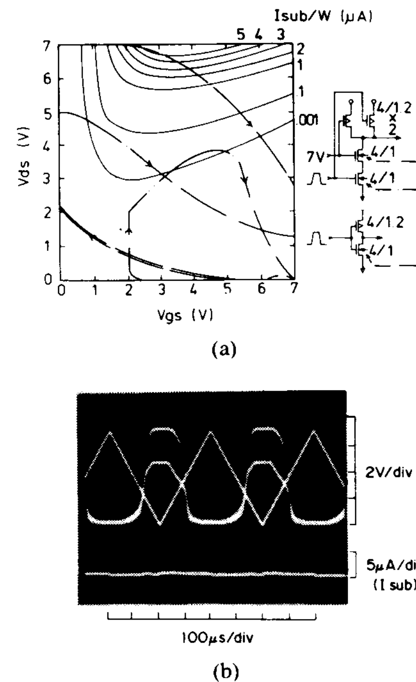


Fig. 7. (a) $V_{gs}-V_{ds}$ trajectories of NAND gate, when the upper transistor turns on first followed by the lower transistor, and (b) measured substrate-current characteristics of the NAND gate. A very small amount of substrate current is generated in this case compared with the inverter case in Fig. 2. t_s is 0.3 ns and load capacitance is 40 fF for (a) and (b).

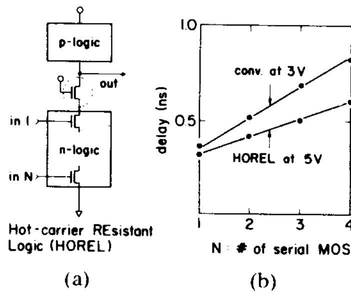


Fig. 8. HOREL: (a) basic configuration, and (b) delay characteristics. The HOREL is characterized by the NOEMI at the top of an n-channel logic tree.

system to achieve higher speed performance over 3-V operation. Since the p-channel MOSFET shows three orders of magnitude better hot-carrier reliability, it is not required to decrease the operation voltage for p-channel logic part. For n-channel logic part, a voltage swing of the gate input is 5 V in the HOREL, which is to be compared with 3 V in the conventional approach.

Since serially connected transistors can be efficiently layed out, silicon area penalty of the inserted MOSFET is less than 30 percent even in an inverter. For NAND gates, penalty is less than 20 percent. In a memory VLSI, area occupation of random logic part is less than 10 percent. Therefore total chip area increase is less than 3 percent, which is considered negligible. In a logic VLSI, random logic part is, say, 20 percent excluding RAM, ROM, wiring, and bonding pad area. Considering more than half of the gates are not simple inverters but complex gates, the total chip increase is about 5 percent. These area estimations are for the case where all the logic part is constructed by using the HOREL. However, the selective use of the NOEMI technology is also effective in VLSI's. Some examples are demonstrated in Sections IV and V.

It is reported [8] that in MOSFET with oxide thinner than 10 nm, hot-carrier generation mode can be changed. However, even for the different generation modes the NOEMI is considered to be effective since the modes show the exponential dependence on V_{ds}.

IV. Static RAM Case

Application to VLSI static RAM [9] is demonstrated in Fig. 9. Simulated duty ratios are shown for heavily loaded bus drivers. The definition of the hot-carrier duty ratio is in the figure. $I_{sub,peak}$ is a peak substrate current at the worst dc bias condition that is usually adopted in a dc measurement for the hot-carrier resistivity of the MOSFET. A more detailed explanation of $I_{sub,peak}$ is in Section II. The most dangerous transistor in terms of hot-carrier generation is a transfer gate of a memory cell. This is because when read "0" operation takes place and a word-line is turning on, the $V_{gs}-V_{ds}$ trajectory for this transistor passes through the point where V_{ds} is V_{cc} and V_{gs} is about the middle of V_{cc}. Applying lower voltage to a bit-line helps to decrease hot substrate current drastically without increasing bit-line delay [10]. Since the duty ratio for the

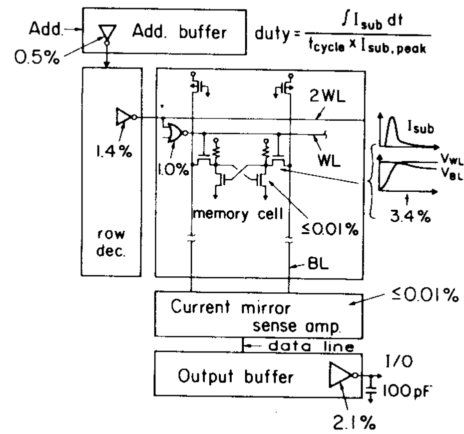


Fig. 9. Hot-carrier duty ratios for 256-kbit CMOS SRAM [9]. The definition of the duty ratio is shown in the figure. The meaning of $I_{sub,peak}$ is explained in Section II. t_{cycle} is assumed to be 46 ns, which is the minimum cycle time of the RAM.

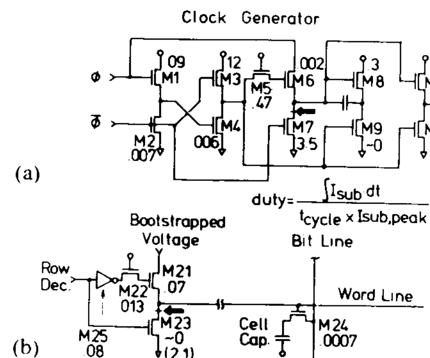


Fig. 10. Hot-carrier duty ratios for dynamic RAM circuits. (a) Widely used E/E type clock generator, and (b) memory-cell peripheral circuits. The meaning of $I_{sub,peak}$ is explained in Section II. Values are in percents. Black arrows indicate the place where NOEMI is required. Duty of $M23$ is 2.3 percent (in the bracket) when the control timing is set so as to discharge the word-line charge through this MOSFET $M23$. The duty is about zero if the word-line is discharged through other MOSFET's. But in this case, there exists some other MOSFET's that discharge the bootstrapped lines and those MOSFET's are in serious condition in terms of hot-carrier generation.

memory-cell transfer gate is almost proportional to word-line delay, aluminum word-line is preferable. On the other hand, driver transistors in a memory-cell flip-flop have a low duty ratio. It is interesting to note that the current-mirror sense amplifier shows a low duty ratio, although the circuit contains direct current path from V_{cc} to V_{ss}. This is because the circuit includes serial connection of nMOSFET's.

V. Dynamic RAM Case

Fig. 10 shows typical DRAM circuits with simulated hot-carrier duty ratios. A cycle time is assumed to be 100 ns in calculating the duty ratio. In a clock generator (Fig. 10(a)), MOSFET $M7$ suffers from a high duty ratio, because its drain is bootstrapped to about $1.5V_{cc}$ and large amounts of hot-carriers are generated at discharging of the high voltage. The same kind of situation occurred in a word-line driver (Fig. 10(b)). When the bootstrapped word-line is discharged through $M23$, the MOSFET is in a serious condition and NOEMI is required at the arrowed

point. It can be said that these MOSFET's are operated under an 8-V system due to the bootstrapped circuit nature, although the supply source voltage is 5 V.

If the NOEMI is applied to the nodes pointed out by a black arrow, the duty ratios are decreased by more than three orders of magnitude. Since the gate of the inserted enhancement MOSFET is biased to V_{cc}, the middle node is charge only up to $V_{cc}-V_{th}$, where V_{th} is a threshold voltage of the inserted MOSFET with a body effect. The speed and area overhead of the inserted MOSFET is negligibly small.

To summarize, certain MOSFET's that discharge bootstrapped nodes generate many more hot carriers than any other transistors. Selective use of NOEMI circuit technology for these highly stressed MOSFET's are effective in reducing the generated hot-carrier amount without degrading operation speed.

It is not effective to use longer thannel devices for these highly stressed MOSFET's. Substrate current of 2.0-μm MOSFET can be reduced only by a factor of three compared with 1.0-μm gate device. This is because the high electric field is determined mainly by junction depth, impurity profiles, and oxide thickness near the drain. Moreover, use of the longer channel device increases an input gate capacitance, which is not the case in the NOEMI.

Another point of interest is about a precharge level of bit-lines. A half-V_{cc} precharge scheme is adopted to simulate the memory array operation. The duty ratio of the cell transfer gate ($M24$) is smaller than in the case of V_{cc} precharge scheme by about three orders of magnitude. This is because the maximum V_{ds} is limited to a half V_{cc} indifferent to the stored data being either "1" or "0" in the half-V_{cc} precharge scheme. If the half-V_{cc} precharge scheme is employed, there is no need to limit the voltage swing of bit-lines by using complex voltage down converter.

VI. Conclusions

Hot-carrier duties in VLSI's depend strongly on $V_{gs}-V_{ds}$ trajectories. High-dutied MOSFET's can be eliminated by carefully designed NAND gates, Clocked CMOS (C^2MOS) gates, the HOREL, and the NOEMI circuit technology. In memories, lowering bit-line precharge voltage is effective, and in special cases in DRAM's a half-V_{cc} precharge scheme is recommended.

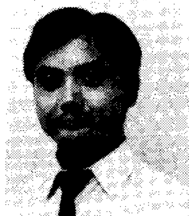
In any case, the amount of generated hot carriers in submicrometer VLSI environment can be reduced by careful design consideration based on circuit physics of $V_{gs}-V_{ds}$ trajectory control. This type of consideration is now beginning and some work has been reported in this course [11], but it will become vital in exploring high-speed and high-density submicrometer VLSI's.

Acknowledgment

The authors wish to thank Dr. Y. Nishi for encouragement throughout the work and M. Isobe, J. Matsunaga, K. Hashimoto, T. Wada, T. Ohtani, K. Tamaru, M. Saito, K. Sawada, M. Kinugawa, and R. Saleh for valuable suggestions and discussions and for supports in computation and measurements.

References

[1] E. Takeda, H. Kume, Yoshimoto, N. Suzuki, S. Asai, and T. Hagiwara, "Hot-carrier effects in submicron VLSI's," in *Proc.1983 Symp. VLSI Tech.*, 1983, p. 104.
[2] F. C. Hsu and K. Y. Chiu, "Hot-electron substrate current generation during switching transistors," in *Proc 1984 Symp. VLSI Tech.*, 1984, p. 86
[3] T. Sakurai, K. Kakumu, and T. Iizuka, "Hot-carrier suppressed VLSI with submicron geometry," *IEEE ISSCC, Dig. Tech. Papers*, Feb. 1985, pp. 272–273.
[4] Y. W. Sing and B. Sudlow, "Modeling and VLSI design constraints of substrate current," in *IDEM-80*, 1980, pp. 732–735.
[5] C. Hu, "Hot-electron effects in MOSFET's," *IEDM-83*, 1983, 176–181.
[6] A. Vladimirescu and S. Liu, "The simulation of MOS integrated circuits using SPICE2," Univ. of Calif. at Berkeley, ERL Memo ERL-M80/7, Oct. 1980.
[7] Y. Suzuki, M. Hagiwara, and K. Odagawa, "Clocked CMOS calculator circuitry," in *IEEE ISSCC, Dig. Tech. Papers*, Feb. 1973, WPM6.3.
[8] E. Takeda, "Hot-carrier effects in submicrometer MOS VLSI's," *Proc. Inst. Elec. Eng.*, vol. 131, pp. 153–162, Oct. 1984.
[9] T. Sakurai, J. Matsunaga, M. Isobe, T. Ohtani, K. Sawada, A. Aono, H. Nozawa, T. Iizuka, and S. Kohyama, "A low power 46-ns 256-kbit CMOS static RAM with dynamic double word line," *IEEE J. Solid State Circuits*, vol. SC-19, no. 5, pp. 578–585, Oct. 1984.
[10] T. Sakurai, K. Sawada, and T. Iizuka, "VLSI-oriented voltage down converter with sub-main configuration," in *Proc. Int. Conf. Solid State Device and Materials* (Late News Abstr.), LC-12-7, 1984, pp. 74–75.
[11] L. Pfennings, W. Mol, J. Bastiaens, and J. Dijk,, "Differential split-level CMOS logic for sub-nanosecond speed," in *IEEE ISSCC, Dig. Tech. Papers*, Feb. 1984, pp. 16–17.



Takayasu Sakurai (S)77–M'78) was born in Tokyo, Japan, on January 10, 1954. He received the B. S., M. S. and Ph.D degrees in electronic engineering from the University of Tokyo, Tokyo, Japan, in 1976, 1978, and 1981, respectively. His Ph.D work is on electronic structures of an $Si-SiO_2$ interface.

In 1981 he joined the Toshiba Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, where he was engaged in the research and development of CMOS dynamic and 64K and 256K static memories. During the development he also worked on modeling of wiring capacitance and delay, new soft-error free memory cells, new memory architectures, and new hot-carrier resistant circuits.

Dr. Sakurai is a member of the Institute of Electronics and Communication Engineers of Japan and Japan Society of Applied Physics.

Kazutaka Nogami was born in Oita, Japan, on May 19, 1959. He received B. E. and M. E. degrees in physical engineering from Tokyo University, Tokyo, Japan, in 1982 and 1984, respectively.

He joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, in 1984, where he has been engaged in research and development of CMOS LSI memory.

Masakazu Kakumu was born in Nagoya, Japan, on January 11, 1956. He received the B. S. and M. S. degrees, both in electrical engineering, from Waseda University, Tokyo, Japan, in 1979 and 1981, respectively.

In 1981, he joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan. His initial research was in silicide materials for application for CMOS VLSI's. Since 1983, he has been engaged in the development of advanced CMOS technologies.

Mr. Kakumu is a member of the Japan Society of Applied Physics and the Institute of Electronics and Communication Engineers of Japan.

Tetsuya Iizuka(M'79) was born in Ibaraki, Japan, on April 17, 1947. He received the B. S. degree in applied physics and the M. S. and Ph.D. degrees in electrical engineering from the University of Tokyo, Tokyo, Japan, in 1970, 1972, and 1975, respectively.

In 1975 he joined the Toshiba Research and Development Center, Toshiba Corporation, Kawasaki, Japan, where he was engaged in research on I^2L and CMOS/SOS devices and in the development of dynamic RAM testing methods. In 1979 he joined a newly organized laboratory, the Semiconductor Device Engineering Laboratory (SDEL), Toshiba Corporation. He developed first generations of 16K, 64K, and 256K CMOS SRAMS's. He also developed a high-speed 64K Bi-CMOS SRAM with bipolar sense amplifiers. He worked on new circuit techniques such as hot-carrier resistant logics. In 1981 he stayed for one year at the Integrated Circuit Laboratory, Hewlett-Packard Laboratories, Palo Alto, CA, as a Visiting Engineer. He studies CMOS latch-up modeling and bird's beak-free isolated MOST analysis. He is serving as the Program Committee Member for the 1985 and 1986 Symposiums on VLSI Technology, and as the Integrated Circuit Subcommittee Member of the 1985 IEDM. He is currently Manager of the Advanced SRAM, EPROM, and EEPROM Design Groups of the SDEL.

Dr. Iizuka is a member of the Institute of Electrical and Communication Engineers of Japan.